

Evolving a NASA Digital Object Identifiers System with Community Engagement

Lalit Wanchoo¹ (lalit.wanchoo@nasa.gov) and Nathan James² (nathan.l.james@nasa.gov),

¹ ADNET Systems, Inc., 7515 Mission Drive, Suite A100, Lanham, MD 20706 , ² Earth Science Data and Information System Project (Code 423), NASA Goddard Space Flight Center, Greenbelt, MD 20771

Introduction

In 2010, NASA's Earth Science Data and Information System (ESDIS) Project began investigating the assignment of unique identifiers to its suite of data products being stewarded at different data centers across the country. This led to the use of Digital Object Identifiers (DOIs) and the development of an automated system for the registration of DOIs.

A key factor in the successful evolution of the DOI registration system has been the incorporation of community input. Over the last three years, ESDIS solicited community input for making the DOI registration process more efficient from three focus groups under NASA's Earth Science Data System Working Group (ESDSWG). These groups were largely composed of DOI submitters and data curators from the 12 data centers serving user communities of various science disciplines. The suggestions from these groups were formulated into recommendations for ESDIS consideration and implementation.

The ESDIS DOI registration system has evolved to be fully functional with over 3,500 publicly accessible DOIs and over 200 DOIs being held in reserve status until the information required for registration is obtained. The goal is to assign DOIs to the entire 7000+ data collections under ESDIS management via its network of discipline-oriented data centers.

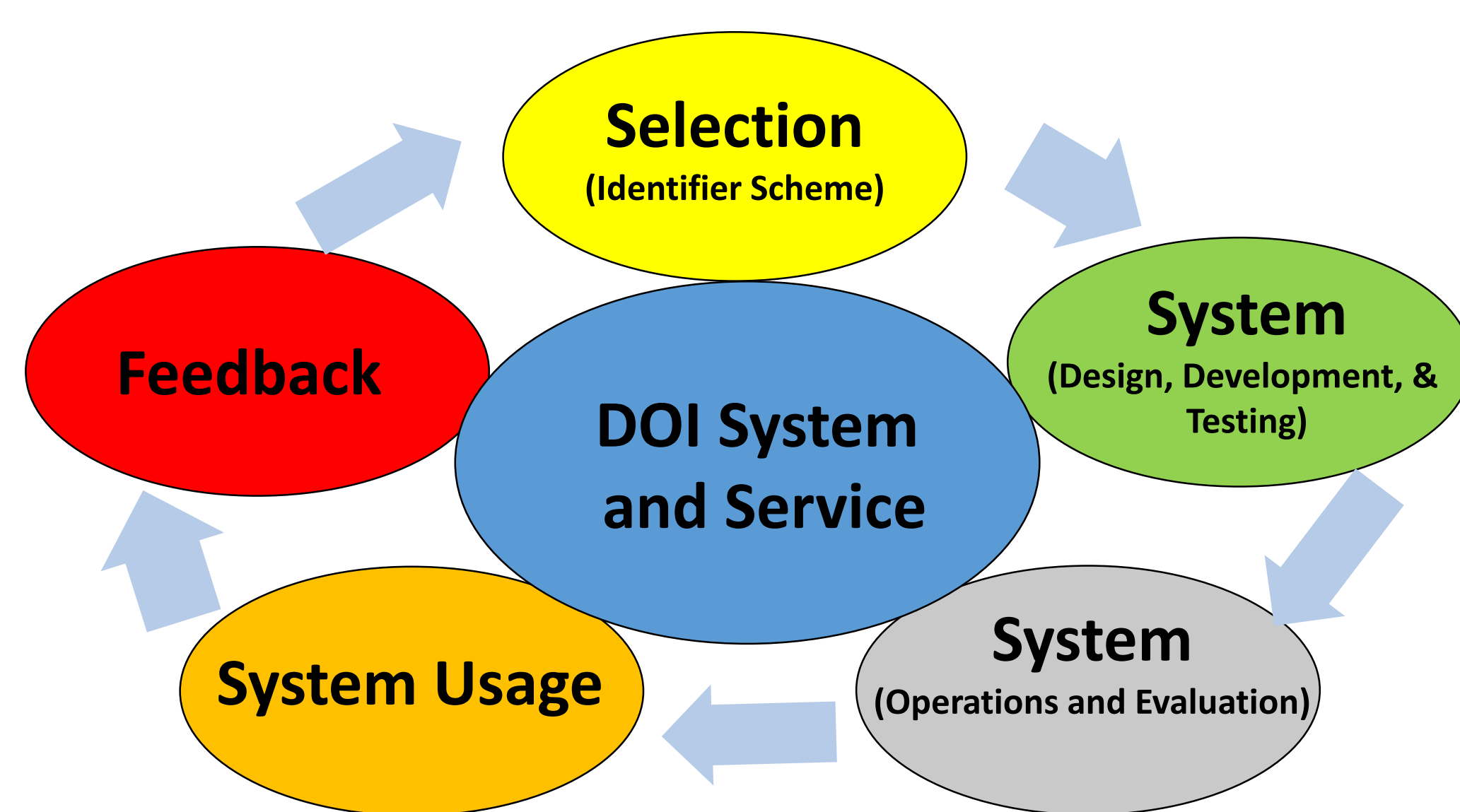
Objective

To demonstrate how the ESDIS DOI system and its processes have evolved over these years based on the recommendations provided by the user community (whether the community members create and manage DOI information or use DOIs in the data citations) The user community is comprised of people with common interests and needs for data identifiers who are actively involved in the creation and usage process. Engagement describes the interactive context wherein the community provides information, evaluates the proposed processes, and provides guidance in the area of identifiers.

Benefits of Community Engagement

Development and implementation of processes, services, and systems that best serve the user community.

Community Engagement Approach



ESDIS established user sub-groups through the Earth Sciences Data Systems Working Group (ESDSWG) to gather input and evaluate the DOI System and Services. These groups addressed each of the identified components of the community engagement approach. Each approach as implemented by ESDIS is described in the following sections:

Selection of Identifier Scheme

ESDIS gave responsibility to a group that included experts from various Distributed Active Archive Centers (DAACs) to evaluate the various available schemes for assigning permanent identifiers to data products distributed by various DAACs. This group evaluated nine applicable schemes as shown in Table 1.

Table 1: Suitability Assessment of Selected Identifier Schemes

Identifier Scheme	Full Name	Unique Identifier	Unique Locator	Citable Locator	Scientifically Unique Identifier
URL/URN/URI	Uniform Resource Identifiers/Names/Locators				
PURL	Persistent Uniform Resource Locators				
XRI	Extensible Resource Identifiers				
Handle	Handle System				
DOI	Digital Object Identifiers				
ARK	Archival Resource Keys				
LSID	Life Science Unique Identifiers				
OID	Object Identifiers				
UUID	Universally Unique Identifiers				

Adapted from Duerr, R. E., et al., 2011. On the utility of identification schemes for digital Earth science data: An assessment and recommendations. *Earth Science Informatics*. DOI 10.1007/s12145-011-0083-6

System Development

ESDIS developed a system that addressed the needs of the community and provided a process that allowed identifiers to be reserved and registered. The operational system is fairly automated and uses a Perl script with an Oracle database for the data. The system functionality provides reports using Oracle Apex.

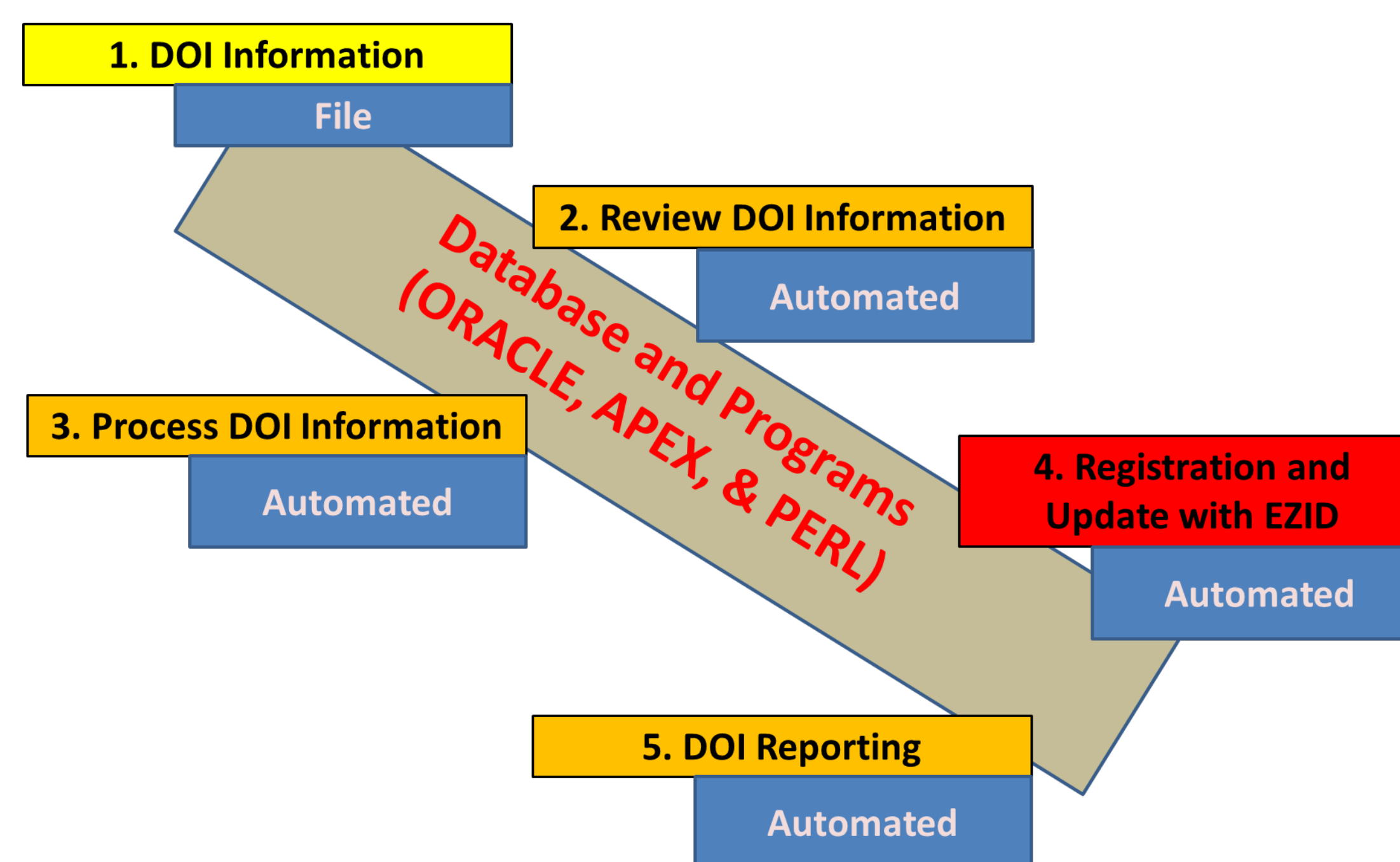


Figure 1: Components of ESDIS DOI Registration and Management System

System Evaluation

The ESDSWG established a DOI Working Group to analyze the approaches and practices for DOI usage across the Earth Observing System Data and Information System (EOSDIS) data centers and DAACs. This group analyzed the processes undertaken by ESDIS in the reserving and registering of DOIs.

ESDIS DOI WIKI URL: <https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS>

In addition, the group evaluated the information being made available to the people who are assigning DOIs to the products. Recommendations from this group and the action taken by ESDIS are shown in Table 2. These recommendations were accepted and implemented.

Table 2: Current Practices Recommendations

Recommendation	Resolution
DOI suffixes should be opaque strings	ESDIS DOI suffix models will include alphanumeric (opaque) strings in
DOIs should be generated by the registry	ESDIS will generate all opaque DOIs with appropriate input from the data providers.
DOIs can be registered using a data center's EZID account or the ESDIS EZID account.	All DOIs for Earth Observing System Data and Information System (EOSDIS) products should be registered using the ESDIS account assigned by EZID unless prior arrangements have already been made.
DOIs should have a corresponding data set landing page.	All ESDIS DOIs must have an accompanying product landing page URL before being registered with EZID. ESDIS can reserve the DOI while the landing page is being developed.
DOIs should be exported to the Global Change Master Directory (GCMD) in the Directory Interchange Format (DIF).	GCMD has already revised the DIF in order to accommodate the DOI information.

Usage and Feedback

To promote consistency, discoverability, and usefulness across NASA, the ESDSWG established three working groups to analyze the implementation and usage of DOIs. These groups recommended the landing page and data citation requirements for the data products that are distributed by EOSDIS. Such recommendations, listed in Tables 3 and 4, were accepted and implemented. The Data Citation group used the Mandatory and Suggested content for citations given in the Data Citation Guidelines for Data Providers and Archives document developed and approved by the Federation of Earth Science Information Partners (ESIP) Federation. These guidelines shall be followed to the greatest extent possible in developing citations for ESDIS data products.

Table 3: Landing Page Recommendations

Landing Page Element	Rationale	Usage
DOI Name	Provides confirmation that the user has arrived at the appropriate landing page.	Required
Data Set Long Name	Assists the users in identifying the data set.	Required
Version Number	Denotes the uniqueness and data production provenance of the data set.	Required
Data Set Description	Provides a more detailed explanation of the data set referenced, so that the user can move from an arbitrary DOI to a more complete understanding of the data.	Required
Data Access	Provides access to the data or notification of how to obtain the data or, where appropriate, that the data are no longer accessible. The official citation for the data set	Required
Data Citation	Identifies the data set lineage through the information contained on the landing pages for other versions of the data sets.	Required
Associated Landing Pages		Recommended

Table 4: Data Citation Recommendations

Citation Attributes	DOI Metadata Attributes
Author	Creator
Release Data	Year
Title	Title
Version Number	(In the Title)
Distributor	Distributor
Access Date	Access Date
Identifier	DOI Name
Example: AIRS Science Team/Joao Teixeira; (2013): Aqua AIRS Level 3 Standard Daily Product using AIRS and AMSU with HSB-V6; NASA Goddard Earth Sciences Data and Information Services Center; May 2014; http://dx.doi.org/10.5067/AQUA/AIRS/DATA302	Example: AIRS Science Team/Joao Teixeira; (2013): Aqua AIRS Level 3 Standard Daily Product using AIRS and AMSU with HSB-V6; NASA Goddard Earth Sciences Data and Information Services Center; May 2014; http://dx.doi.org/10.5067/AQUA/AIRS/DATA302

Conclusion

The ESDIS DOI System and its registration process has been primarily developed based on the input from the EOSDIS community. This successful system, which has been in operation for the last four years, has processed over 5,000 DOI requests from 11 data providers (see Figure 2). Currently, over 4,000 DOIs have been registered.

Acknowledgments

Lalit Wanchoo's contributions to this study were funded through the Science and Exploration Data Analysis (SESDA III) GSFC NASA Contract No: NNG12PL17C and Nathan James's contributions to this study were made as a part of his employment by NASA.

Figure 2: Monthly DOI Requests Received by ESDIS for Registration

